DOCUMENT RESUME

ED 406 447 TM 026 427

AUTHOR Whitaker, Jean S.
TITLE Use of Stepwise Methodology in Discriminant Analysis.
PUB DATE 23 Jan 97
NOTE 33p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23-25, 1997).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Discriminant Analysis; Prediction; Regression (Statistics); *Research Problems; Sampling; Selection
IDENTIFIERS Cross Validation; *Stepwise Regression

ABSTRACT

The use of stepwise methodologies has been sharply criticized by several researchers, yet their popularity, especially in educational and psychological research, continues unabated. Stepwise methods have been considered particularly well suited for use in regression and discriminant analyses, but their use in discriminant analysis (predictive discriminant analysis and descriptive discriminant analysis) has not been the direct focus of as much written commentary. This discussion considers several problems associated with the use of stepwise methods. These include: (1) incorrect degrees of freedom associated calculated by computer packages; (2) sampling error capitalization; and (3) failure to select the best subset of predictors. Although they offer the promise of assisting researchers with tasks such as variable selection and variable ordering, this promise is almost always unfulfilled. Some alternatives to the use of stepwise methods are discussed, including manually correcting degrees of freedom, cross-validation procedures, and all-possible subsets analyses. (Contains 6 tables and 45 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running head: STEPWISE METHODOLOGY USE IN DISCRIMINANT ANALYSIS

ED 406 447

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
JEAN S. WHITAKER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Use of Stepwise Methodology in Discriminant Analysis

Jean S. Whitaker

Texas A&M University 77843-4225

BEST COPY AVAILABLE

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 23, 1997.

TM026427

Abstract

The use of stepwise methodologies has been sharply criticized by several researchers, yet their popularity, especially in educational and psychological research, continues unabated. Stepwise methods have been considered particularly well suited for use in regression and discriminant analyses, however their use in discriminant analysis (predictive discriminant analysis and descriptive discriminant analysis) has not been the direct focus of as much written commentary. Therefore, predictive discriminant analysis and descriptive discriminant analysis are discussed in general, and then their relevance with respect to stepwise techniques is examined. There are several problems associated with the use of stepwise methods. Stepwise methods hold out the promise of assisting researchers with such important tasks as variable selection and variable ordering. However, the promise is almost always unfulfilled and researchers are cautioned against using stepwise methodologies. Some alternatives to the present use of stepwise methods are discussed.

Stepwise Methodology in Discriminant Analysis

Huberty (1989) stated that discriminant analysis (DA) includes a set of response variables and a set of one or more grouping or nominally scaled variables. Klecka (1980, p. 8) outlined the basic prerequisites for conducting a discriminant analysis, i.e., "two or more groups exist which we presume differ on several variables and that those variables can be measured at the interval or ratio level." There is no limit to the types of variables that can be employed, but problems with interpretation may result (Nunnally & Bernstein, 1994). Klecka (1980, p. 8) noted that "discriminant analysis will then help us analyze the differences between the groups and/or provide us with a means to assign (classify) any case into the groups which it most closely resembles."

As Huberty's description and Klecka's prerequisites in the above paragraph imply, discriminant analysis has two sets of techniques based on the purpose of the analysis, i.e., predictive discriminant analysis and descriptive discriminant analysis. "When groups of units are known in advance and the purpose of the research is either to describe *group* differences [DDA] or to predict group membership [PDA] on the basis of response variable measures, discriminant analysis techniques are appropriate" (Huberty, 1994). Alternatively, Stevens (1996) described the distinction between PDA and DDA in the following way: "in predictive discriminant analysis the focus is on classifying subjects into one of several groups, whereas in descriptive discriminant analysis the focus is on revealing major differences among the groups" (Stevens, 1996).

Discriminant analysis has been described by some researchers as similar to multiple regression (MR) analysis (Gall, Borg, & Gall, 1996) inasmuch as it is an adaptation of regression analysis techniques (Kachigan, 1986). In fact, anyone who is

familiar with the basic goals and techniques of multiple regression can easily understand the association between multiple regression and discriminant analysis. Unfortunately, as Kachigan (1986, p. 375) pointed out, DA is sometimes used in instances where regression analysis is a "more appropriate and powerful technique." In these cases the continuous criterion variable is dichotomized in order to lend the analysis to DA procedures. This practice of squandering variance information, by dichotomization or polychotomization of continuous variables, has been strongly criticized in the literature (Kerlinger, 1986; Thompson, 1994).

Despite the close association between DA and MR, it is important to note that some researchers have recognized that all parametric procedures can be derived from the same linear model which involves the use of least squares weights (Cohen, 1968; Knapp, 1978). As Knapp (1978) noted, "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis, which is the general procedure for investing differences between two sets of variables" (p. 410). Thompson (1988) pointed out that every parametric procedure involves the creation of a synthetic score(s) for each individual on some latent construct. In discriminant analysis the synthetic scores are the discriminant scores created with the discriminant function coefficients (Pedhazur, 1982).

A researcher must make choices about the variables that will be involved in an analysis. Oftentimes, the researcher may want (a) to select a subset of variables from the original set or (b) to determine the relative importance of the set of variables even if no variables are to eliminated. Some researchers erroneously believe that stepwise methods can be used to accomplish either of these tasks (Huberty, 1989). Also, stepwise methods

can be used with either PDA or DDA, but their application in PDA in particular is rarely appropriate (Huberty & Barton, 1989).

Several researchers (Huberty, 1989; Huberty, 1994; and Thompson, 1989) have noted the common use of stepwise analyses. According to Thompson (1989, p146), "stepwise analytic methods may be among the most popular research practices employed in both substantive and validity research." However, some of these same researchers as well as others (cf. Snyder, 1991) have advanced strong arguments against the use of stepwise methodologies.

A discussion of the problems associated with stepwise methodologies in discriminant analysis is best understood with a basic understanding of discriminant analysis itself. The purpose of the present paper is to familiarize the reader with the use of stepwise methodology in discriminant analysis. Therefore, a brief history of DA and a description of discriminant analysis is offered first. Second, clarification of the use of stepwise techniques in both PDA and DDA, is presented. Third, stepwise methodologies, as applied to DA, and the inherent problems in their use are discussed. Last, a number of alternative suggestions to the use of stepwise procedures are offered.

Discriminant Analysis

History

The ideas associated with discriminant analysis can be traced back to the 1920s and work completed by the English statistician Karl Pearson, and others, on intergroup distances, e.g., coefficient of racial likeness (CRL), (Huberty, 1994). In the 1930s R. A. Fisher translated multivariate intergroup distance into a linear combination of variables to aid in intergroup discrimination. Methodologists from Harvard University contributed

much to the interest in application of discriminant analysis in education and psychology in the 1950s and 1960s (Huberty, 1994). Klecka (1980) provided several historical references that deal mostly with early applications of DA.

The two types of discriminant analysis, i.e., PDA and DDA, have different histories of development. According to Huberty (1994), "discriminant analysis for the first three or four decades focused on the prediction of group membership," PDA, whereas DDA usage did not appear until the 1960s and "its use has been very limited in applied research settings over the past two decades." PDA and DDA are multivariate analyses that have important differences in both their general application and when used in conjunction with stepwise methodology.

The following sections on descriptive discriminant analysis and predictive discriminant analysis are deliberately limited as regards technical and mathematical descriptions. The reader is encouraged to consult the numerous texts on DA referred to by Huberty (1994, pp. 25-26) and Klecka (1980, pp. 14-15) for a more technical treatment of the subject. In addition, many texts on multivariate data analysis have sections or chapters on discriminant analysis; however, some of these texts, especially earlier ones, do not make clear distinctions between PDA and DDA.

Predictive Discriminant Analysis

Predictive discriminant analysis (PDA), or "classification" as it is sometimes called, generally includes "a set of predictor variables and one criterion variable, the latter being a grouping variable with two or more levels, that is, there are two or more groups" (Huberty & Barton, 1989, p. 158). Predictive discriminant analysis is similar to multiple regression analysis except that PDA is used when the criterion variable is categorical and

nominally scaled. As in multiple regression, in PDA a set of rules is formulated which consists of as many linear combinations of predictors as there are categories, or groups (Huberty, 1994). The equation in PDA uses a person's scores on the predictor variables to predict the category to which the individual belongs (Gall, Borg, & Gall, 1996, p. 441).

For example, a school district might be interested in predicting which pre-kindergarten students are likely to have difficulty learning to read by second grade. A prediction rule would be generated using such predictors as scores on a kindergarten readiness test, ratings on age at which developmental milestones were reached, family socio-economic status, and gender. Predictor weights for two linear combinations, one associated with each group, are determined (Huberty, 1994). Two probabilities of group membership can be calculated for subsequent students based on the two linear combinations; the student is assigned to the group with the larger linear combination score.

In predictive discriminant analysis each object will have a single score on the discriminant function in place of its scores on the various predictor variables. At the same time a cutoff score will be determined such that when the criterion groups are compared with respect to the discriminant scores the errors of classification are minimized (Kachigan, 1986, p. 365). Table 3 provides an example of a classification table used to report results from an application of a prediction rule. This heuristic provides information about the accuracy of the prediction rule, i.e., "the hit rates" or correct classifications. The overall percentage of correct classifications is 83.3%. The percentage of correct classifications must be judged against chance probabilities. Are our

results better than chance? Yes, chance probabilities here would result in a 56% accurate classification rate.

Descriptive Discriminant Analysis

DDA includes a collection of techniques involving two or more criterion variables and a set of one or more grouping variables, each with two or more levels, whose effects are assessed through MANOVA. "Whereas in predictive discriminant analysis (PDA) the multiple response variables play the role of predictor variables, in descriptive discriminant analysis (DDA) they are viewed as outcome variables and the grouping variable(s) as the explanatory variable(s). That is, the roles of the two types of variables involved in a multivariate, multigroup setting in DDA are reversed from the roles in PDA" (Huberty, 1994, p. 30). In DDA the total "between-groups" association in MANOVA is broken down into additive pieces through the use of uncorrelated linear combinations of the original variables (discriminant functions) (Stevens, 1996, p. 261).

Insert Table 1 About Here

According to Kerlinger & Pedhazur (1973, p. 337) "the discriminant function is a regression equation with a dependent variable that represents group membership." The aforementioned relationship between multiple regression and descriptive discriminant analysis is clearly illustrated in the two-group, or dichotomous grouping variable case, i.e., regression and DDA yield the same results. As can be seen from the heuristic example in Table 1, lambda at a given step equals 1 - R^2 and, conversely, R^2 equals 1 -

lambda. Once the analysis changes to a DDA with more than 2-groups, the calculations become more complex and are no longer directly analogous to regression results.

Insert Table 2 About Here

Again, Table 2 shows the relationship between DDA structure coefficients and regression structure coefficients for the above mentioned case. Although values are not identical and are arbitrarily scaled in the opposite direction, their relative magnitudes within each column are the same.

Stevens (1996) pointed out that DA makes descriptions parsimonious because 5 groups can be compared on 10 variables, for example, where the groups differ mainly on only two major dimensions (discriminant functions). In DDA linear combinations are used to distinguish groups. If k is the number of groups and p is the number of dependent variables, then the number of possible discriminant functions is the minimum of p and (k - 1) (Stevens, 1996, p. 263).

Again, DDA is associated with MANOVA. When the results of the omnibus MANOVA effects has been shown to be generalizable, DDA can be used to describe and interpret these effects. The linear composites (linear discriminant functions, LDFs) can be used to identify outcome variable "constructs (or latent variables) that underlie the group differences, that is, that underlie the grouping variable effect" (Huberty, 1994, p. 206).

Huberty (1994) stated that "the predominant method of identifying latent constructs in multivariate analyses--this includes factor analysis and canonical

correlation--is to examine correlations between linear composite scores and scores on the individual variables in the composite. These LDF-variable correlations are often called *structure r's*" (p. 209). Interpretation of the latent construct underlying these functions is largely based on these structure r's (Huberty & Barton, 1989).

Summary of PDA and DDA

The present paper has outlined the differences between DDA and PDA. Aside from the differences in purpose, variable roles, and two aspects of DA, the sampling designs may be also be different (Huberty, 1989, p. 159). One of the most important differences for the researcher is that of purpose. Although Huberty and Barton (1989) noted that some studies report both a PDA and a DDA for the same data, it is unlikely that both types of DA are relevant to the research question(s). However, Klecka (1980) presented a case that examined Senatorial factions and utilized both procedures. But, as Huberty and Barton (1989, p. 166) aptly stated,

> the purposes of the two analyses are different, the roles of the two sets of variables in each analysis are *reversed*, and the techniques in the two analyses are different. There is, perhaps, some feasibility to the "mixing" of DDA and PDA for purposes of corroboration of results. But, generally, research questions are of the descriptive type or of the predictive type; only seldom would both types of questions be addressed in a given research situation.

PDA is appropriate when the researcher is interested in assigning units (individuals) to groups based on composite scores on several predictor variables, (i.e., LCFs). The accuracy of such prediction can be assessed by examining "hit rates" as against chance, for example. The most basic question answered by PDA is "given the

individuals scores on several predictor variables, which group represents their true membership group?" Again, the focus of PDA is prediction and the accuracy of hit rates. As Huberty and Barton (1989) noted with respect to PDA, "One is basically interested in determining a classification rule and assessing its accuracy..."

Huberty and Barton (1989) noted that some authors contend that a MANOVA should be conducted prior to considering a PDA or DDA, however, they cite three reasons for not doing MANOVA within a PDA context. On the other hand, DDA is statistically intertwined with MANOVA results. When MANOVA results suggest that there are group differences, i.e., true effects, then DDA can be used as a post hoc method to assess the predictor variables that best explain this group separation.

Some researchers incorrectly use a series of post hoc ANOVAs to investigate statistically significant MANOVA effects, but this is inappropriate since univariate methods can <u>not</u> be used to explore multivariate effects. However, DDA is a multivariate method, and DDA can indeed be quite useful as a post method to employ following a MANOVA (Thompson, 1994c).

Stepwise Methodologies

<u>History and Introduction</u>

According to Huberty (1989), "stepwise analysis is believed to have been first advanced by Efroymson (1960), and is fully described by Draper and Smith (1981, chap. 6) and Jennrich (1977a, 1977b)." Several variants of stepwise methods are available through the statistical packages, e.g., forward selection, backward elimination, forward stepwise, and backward stepwise analysis. However, the default settings usually result in

through the statistical packages, e.g., forward selection, backward elimination, forward stepwise, and backward stepwise analysis. However, the default settings usually result in a forward selection analysis (Huberty, 1994, p. 261). Stepwise methodologies have enjoyed popular usage, especially in educational and psychological research settings. Huberty (1994) noted the widespread use of stepwise methods in empirically based journal articles. Thompson (1989) suggested that "stepwise analytic methods may be among the most popular research practices employed in both substantive and validity research" (p. 146).

Researchers erroneously use stepwise methods to evaluate the relative importance of variables in a particular study or to choose variables to retain for future analyses. However, a number of researchers have cautioned against using stepwise methodologies because they fail to achieve the aforementioned two purposes, namely, to evaluate variable importance or to select variables. In addition, there are problems associated with stepwise methodologies in a variety of statistical contexts. The problems with stepwise methods described below are just as relevant within a univariate context, such as regression, as they are in any multivariate case (Moore, 1996).

Huberty (1989, p. 43) stated that three popular computer software packages, i.e., BMDP, SAS, and SPSS, include programs to conduct a "stepwise multiple regression analysis" and a "stepwise discriminant analysis." According to Stevens (1973; as cited in Huberty, 1989, p. 43), "although regression analysis and discriminant analysis problems are, without a doubt, the most popular contexts for the use of step-type computational algorithms, these approaches have also been suggested in multivariate

Problems Inherent in the Use of Stepwise Methodologies

Several researchers (Huberty, 1989, 1994; Snyder, 1991; Thompson, 1989, 1995) have highlighted three basic problems inherent in the use of stepwise methodologies, i.e., incorrect degrees of freedom, sampling error capitalization, and failure to select the best subset of variables of a given size, and they presented scathing criticisms of applications of these techniques. Although these problems are germane to stepwise methodologies in certain univariate cases, e.g., regression, and some other multivariate analyses, the present discussion is confined primarily to discriminant analysis.

Incorrect Degrees of Freedom

First, as noted above, incorrect degrees of freedom are used in the calculation of statistical tests for discriminant function analysis by most computer packages that employ stepwise methods. Although some researchers have challenged traditional interpretations of statistical significance testing (Carver, 1978; Cronbach, 1975; Cohen, 1990, 1994; Meehl, 1990; Shaver, 1993; Thompson, 1993), they are still part of many analyses. When incorrect degrees of freedom are used the results of statistical tests of significance are systematically biased in favor of spuriously high statistical significance (Thompson, 1989). Students and researchers should be cautioned against interpreting potentially fallible results commonly generated by computer packages.

Thompson (1995) remarked that "degrees of freedom in statistical analyses reflect the number of unique pieces of information present for a given research situation. These degrees of freedom constrain the number of inquiries we may direct at our data and are the currency we spend in analysis" (p. 526). In any computerized stepwise procedure the pre-set degrees of freedom are "one" for each variable included in the analysis.

Thompson (1994) drew the analogy of the pre-set degrees of freedom as coins that we can spend to explore out data, or rather, we are charged one degree of freedom for every predictor variable used. However, at each step all predictors from the original variable set are considered for inclusion. So, at each step the correct number of degrees of freedom should be the same as the total number of variables from the predictor set. If the original number of predictor variables was ten than the correct "charge" is ten.

In the statistical test of significance, there are three calculations for degrees of freedom, i.e., explained, unexplained, and total. The computer packages calculate the correct "total" degrees of freedom (n-1). In regression the "explained" degrees of freedom are erroneously entered as the number of predictor variables (i.e., pv). Therefore, in regression the degrees of freedom "unexplained" (1-pv) are necessarily computed incorrectly (Thompson, 1995). Thompson provided a clear illustration of this type of error within a regression context in that same journal article.

Insert Table 4 About Here

Thompson's example involves data from 101 subjects on dependent variable, ("Y"), and 50 predictor variables. As can be seen from Table 4, the degrees of freedom computed by the computer packages (Analysis 1) yield a statistically significant (α=.05) result. However, the correct degrees of freedom are given in Analysis 2. As Thompson noted, "If the five entered predictor variables had been randomly selected, an explained degree of freedom of 5 might be arguably correct" (p. 527). The error is built into computer programs that do discriminant analyses.

Sampling Error Capitalization

Second, stepwise techniques often capitalize greatly on even small amounts of sampling error and, thereby, reduce the generalizability of results (Davidson, 1988; Snyder, 1991; Thompson, 1995). Lack of generalizability pertains directly to the question of replicability. Replicability of results is important in research endeavors.

This capitalization on sampling error is possible because of the way in which stepwise analyses (forward stepwise analyses) choose variables. In a stepwise analysis variables are entered one at a time within the context of previously entered variables, in a one-at-a-time fashion. In other words, the first variable chosen is the one with the most variance explained, the second one chosen, in the second step, is that variable that has the next best amount of explained variance that does not overlap with the first variable chosen, i.e., unique variance. It is conceivably that two variables, call them V_1 and V_2, may have very similar explanatory ability, with variance accounted for that is infinitesimally different from each other. In fact, these differences may be due only to sampling error and represent little, if any, true difference.

The stepwise procedure will choose the variable with the most explanatory power in this particular sample in the first step and the second variable, say V3, chosen in the second step, based on its ability to account for new and unique variance, in this sample and given exactly this set of predictors. The packages do not provide interpretation. If the variable that was ignored in the first step, V2 was more practical or economical, or if its true population effect was even larger, V2 would still be ignored. As Thompson (1995) suggested, it is possible that otherwise worthy variables are often excluded from the analysis altogether and assumed to have no explanatory or predictive potential.

For example, see Tables 5 and 6. Table 5 presents standardized canonical discriminant function coefficients and Table 6 presents a structure matrix from a stepwise discriminant function analysis. On the two functions listed in Table 5, it appears that variable Y_4 provides the greatest amount of explanatory power on the first function and, correspondingly, variable Y_1 on the second function. If we based our interpretation of the results solely on the information in Table 5, we would form an erroneous conclusion. However, Table 6 describes a different picture of the potential explanatory power of individual variables.

Insert Tables 5 and 6 About Here

Numerous researchers have pointed out the benefits of interpreting structure type coefficients in both regression analysis (Bowling, 1993; Daniel, 1990; Perry, 1990; Thompson, 1992; Thompson & Borrello, 1985) and discriminant analysis (Pedhazur, 1982). In the present example, Table 6, the structure matrix reveals that variable Y_3, on function one, and variable Y_2, on function two, also contain much explanatory ability, or ability to account for variance. Therefore, the explanatory ability of variables Y_4 and Y_3 on the first function, and variables Y_1 and Y_2 on the second function, are very similar. The differences between Y_4 and Y_3, or between Y_1 and Y_2, may be due to sampling error.

An important aspect of any scientific endeavor is replication. It is unlikely that these small differences, which may be due to sampling error, will replicate. It is conceivable that in future studies variables Y_2 and Y_3 will receive credit for explanatory ability that helps differentiate the groups on Functions I and II, respectively.

Failure to Select the Best Subset of Variables

Third, the fact that stepwise methods do not identify the best predictor set of a given size is also problematic. Stepwise methods do not necessarily identify the best predictor set of a given size (Huberty, 1994; Thompson, 1995), even for the sample data being analyzed. The true best set (a) may yield considerably higher effect sizes and (b) may even include none of the variables selected by the stepwise algorithm.

The reasons why stepwise methods are typically used, i. e., variable selection and variable ordering, may not be accurate due to the potential of these methods to capitalize on small amounts of sampling error. The problem of variable ordering was outlined in the previous section. Variable selection may be important when the original variable set needs to be reduced for a particular reason. Unfortunately, as several researchers have demonstrated (Snyder, 1991; Thompson 1989, 1995), stepwise methodologies are not accurate for either univariate or multivariate purposes.

Thompson (1995), using a stepwise regression example, described how stepwise procedures do not select the best set of predictor variables of size q. For example, five predictors entered in five steps of forward entry will not typically answer the question "What is the best set of q = 5 predictors?"

Stepwise Methods in DDA and PDA

Huberty and Wisenbaker (1992) indicated that the aforementioned computer programs (BMDP, SAS, and SPSS) provide most of the quantitative information needed to interpret a PDA or a DDA. In the computer programs discussed earlier in this paper, BMDP, SAS, and SPSS, PDA is used to classifying subjects into one of several groups,

whereas DDA is designed to reveal major differences among groups. Stepwise methodologies can be used with either form of DA, but is typically used in a DDA context. The major differences between DDA and PDA are "revealed through the discriminant functions (Stevens, 1996). Three statistical packages, BMDP, SAS, and SPSS all perform a stepwise discriminant analysis (also stepwise regression analysis). Huberty (1994, p. 261) stated that "when it is claimed that a "stepwise ____ analysis" was run, more likely than not it was a forward stepwise analysis using default values for variable delection, which usually simply results in a forward analysis."

Stepwise Methods in DDA

The statistical packages named above have stepwise discriminant analysis programs with built-in criteria for stepping that relate to group separation. In a forward analysis, variables are selected at each step such that group separation is increased the most. Therefore, inherent problems aside, stepwise methods would appear to be more appropriate in a MANOVA/DDA context where group separation is the focus of the discriminant analysis (Huberty, 1994, p. 261). Within this context, methods that increase the separation of groups by providing information about the importance of variables, an erroneous enticement offered by stepwise methodologies, would be valuable

Stepwise Methods in PDA

Stepwise methods in a PDA context, where group membership prediction is the point of the analysis, would only be considered in "very restrictive situations" (Huberty, 1994, p. 261). According to Huberty (1989, p. 166), "... it has not been shown that package stepwise results are relevant for a predictive discriminant analysis."

Since PDA is concerned with hit rates and accuracy of classification, any reasonable PDA stepwise procedures must focus on maximizing hit rates. The widely used computer packages do not have stepwise algorithms that do this.

Better Alternatives to Stepwise Methodologies

The problems inherent with stepwise methodologies as outlined above are serious. Researchers such as Huberty (1989) do not "espouse the use of stepwise analyses" (p. 48). Huberty recognized that reducing the number of variables is sometimes warranted, as a preliminary analysis, e.g., to reduce the number of response variables to a manageable size. He suggested that variables be discarded when they do not provide predictive validity, for example, those that have contributed little to predictive validity in previous studies, variables highly correlated with other variables, and variables that are judged not relevant to the present study.

The problem of incorrect degrees of freedom in statistical tests of significance could be addressed directly by the researcher by changing the values to the correct ones and recalculating the F statistics. In other words, the problems with degrees of freedom in the computer packages can be remedied by individual researchers before they interpret their results. The incorrect degrees of freedom calculated by the computer packages can simply be corrected by hand.

There are methods to determine the best subset of variables of size q. There are different ways to address the problem; however, perhaps the best solution is to use an "all-possible-subsets" approach (Huberty, 1989; Thompson, 1995). More specifically, researchers could conduct an all possible subsets of each size in order to determine the best subset of any given size. Computer programs are available that do this painlessly.

Conclusion

Despite the adamancy with which certain scholars caution the unwary researcher against using stepwise methods, their use continues unabated. The problems associated with stepwise methods, i.e., incorrect degrees of freedom calculated by computer packages, sampling error capitalization, and failure to select the best subset of predictors, are significant. Some of the alternatives to address these problems, such as manually correcting degrees of freedom, cross-validation procedures, and all-possible subsets analyses, have been forwarded in the present paper. Perhaps the best alternative for researchers is to remember that computer packages do what they are programmed to do, and do not provide interpretation of results. It is the researcher who must design the study and choose the statistical procedures that will allow him or her the greatest opportunity for making sense of the results on the computer printouts.

References

Bowling, J. (1993, November). The importance of structure coefficients as against beta weights: Comments with examples from the counseling psychology literature. Paper presented at the annual meeting of Mid-South Education Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 364 606)

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443. (48, 378-399)

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Cooley, W. W., & Lohnes, P. R. (1971). Multivariate data analysis. New York: John Wiley & Sons.

Dalgleish, L. I. (in press). Discriminant analysis: Statistical inference using the jackknife and bootstrap procedures. Psychological Bulletin.

Daniel, L. G. (1990, January). The use of structure coefficients in multivariate educational research: A heuristic example. Paper presented at the annual meeting of the

Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 451)

Davidson, B. M. (1988, November). The case against using stepwise research methods. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville. (ERIC Document Reproduction Service No. ED 303 507)

Draper, N. R., & Smith, H. (1981). Applied regression analysis. New York: Wiley.

Efroymson, M. A. (1960). Multiple regression analysis. In A. Ralston & H. S. Wilf (Eds.). Mathematical methods for digital computers (pp.191-203). New York: Wiley.

Fisher, R. A. (1940). The precision of discriminant functions. Annuals of Eugenics, 10, 422-429.

Friedrich, K. R. (1991, April). The importance of structure coefficients in parametric analysis. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 330 725)

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: an introduction (6th ed.). New York: Longman Publishers USA.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95, 156-171.

Huberty, C. J. (1989). Problems with stepwise methods: Better alternatives. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J. (1994). Applied discriminant analysis. New York: Wiley and Sons.

Huberty, C. J., & Barton, R. M. (1989). An introduction to discriminant analysis. Measurement and Evaluation in Counseling and Development, 22, 158-168.

Huberty, C. J., & Wisenbaker, J. M. (1992). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 43-70). Greenwich, CT: JAI Press.

Jennrich, R. I. (1977a). Stepwise regression. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.). Statistical methods for digital computers (Vol. 3, pp. 58-75). New York: Wiley.

Jennrich, R. I. (1977b). Stepwise discriminant analysis. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.). Statistical methods for digital computers (Vol. 3, pp. 76-96). New York: Wiley.

Kachigan, S. K. (1986). Statistical analysis. New York: Radius Press.

Kerlinger, F. N. (1986). Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods (2nd ed.). New York: Radius Press.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Klecka, W. R. (1980). Discriminant analysis. Beverly Hills, CA: Sage.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. Psychological Reports, 66, 195-244.

Moore, J. D. (1996, January). Stepwise methods are as bad in discriminant analysis as they are anywhere else. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill, Inc.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart, & Winston.

Perry, L. (1990, January). The use of structure coefficients in regression research. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 448)

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. Journal of Experimental Education, 61, 293-316.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 99-105). Greenwich, CT: JAI Press.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: SAGE.

Thompson, B. (1988, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1989). Why won't stepwise methods die? Measurement and Evaluation in Counseling and Development, 21 (4), 146-148.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897)

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. Journal of Experimental research, 61, 361-377.

Thompson, B. (1994a, April). Common methodological mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 368 71)

Thompson, B. (1994b). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62(2), 157-176.

Thompson, B. (1994c, February). Why multivariate methods are usually vital in research: Some basic concepts. Paper presented as a Featured Speaker at the biennial meeting of the Southwestern Society for Research in Human Development, Austin, TX. (ERIC Document Reproduction Service No. ED 367 687)

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. Educational and Psychological Measurement, 55(4), 525-534.

Thompson, B. (in press). Canonical correlation analysis: Basic concepts and some recommended interpretation practices. In L. Grimm & P. Yarnold (Eds.), Reading and understanding multivariate statistics (Vol. 2). Washington, DC: American Psychological Association.

Thompson, B. & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Table 1.

Comparison of DDA (2-groups case) and Regression Results

Step Entered	Vars in	Wilks' Lambda	R^2
1	V_1	0.15730	.84270
2	V_4	0.13390	.86610
3	V_3	0.11914	.88086
4	V_2	0.10838	.89162

Table 2.

Comparison of DDA (2-group case) and Regression Results

Variable	DDA struc	r with Y	Mult R	struc r
V_1	0.80695	-0.9180	0.94426	-0.97219
V_2	-0.31826	0.6742	0.94426	0.71400
V_3	0.33544	-0.6933	0.94426	-0.73423
V_4	-0.36746	0.7254	0.94426	0.76822

Table 3.

Example of a Two-Group Classification Table for 300 6th Graders

	Actual Group		
Predicted Group	Non-drop out	Drop-out	Totals
Drop-outs	25 (errors)	75	100
Non-drop outs	175	25 (errors)	200
Totals	200	100	300

Table 4.

Hypothetical Five-Step Regression Model with 101 Subjects and 50 Predictor Variables

Analysis	Source	SOS	df	MS	Fcalc	Fcrit	R^2
1	Explained	20	5	4.0000	4.75	4.41	20.00%
	Unexplained	80	95	0.8421			
	Total	100	100				
2	Explained	20	50	0.4000	0.25	—[a]	20.00%
	Unexplained	80	50	1.6000			
	Total	100	100				

a. Because $F_{critical}$ at infinite and infinite degrees of freedom equals 1, an $F_{calculated}$ less than 1 cannot be statistically significant.

(Reprinted from Thompson, 1995)

Table 5.

Structure Matrix

Variable	Function 1	Function 2
Y_1	.0977	.7432
Y_2	.2304	.0940
Y_3	.3325	.0085
Y_4	.6989	.1121

Table 6.

Structure Matrix

Variable	Function 1	Function 2
Y_1	-.1005	.8942
Y_2	.3317	.8845
Y_3	.9325	-.0167
Y_4	.9276	.1985